



Okasha, S. (2011). Experiment, observation, and the confirmation of laws. *Analysis*, 71(2), 222 - 232. <https://doi.org/10.1093/analys/anr014>

Peer reviewed version

Link to published version (if available):
[10.1093/analys/anr014](https://doi.org/10.1093/analys/anr014)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in *Analysis* following peer review. The definitive publisher-authenticated version: Samir Okasha. Experiment, observation and the confirmation of laws. *Analysis* (2011), 71 (2), 222-232 is available online at <http://dx.doi.org/10.1093/analys/anr014>

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Experiment, Observation, and the Confirmation of Laws

It is customary to distinguish experimental from purely observational sciences. The former include physics and molecular biology, the latter astronomy and paleontology. Experiments involve actively intervening in the course of nature, as opposed to observing events that would have happened anyway. When a molecular biologist inserts viral DNA into a bacterium in his laboratory, this is an experiment; but when an astronomer points his telescope at the heavens, this is an observation. Without the biologist's handiwork the bacterium would never have contained foreign DNA; but the planets would have continued orbiting the sun whether or not the astronomer had directed his telescope skyward.

The observational/experimental distinction would probably be difficult to make precise¹, as the notion of an 'intervention' is not easily defined, but it is intuitively fairly clear, and is frequently invoked by scientists and historians of science. Experimentation, or 'putting questions to nature', is often cited as a hallmark of the modern scientific method, something that permitted the enormous advances of the last 350 years. And it is sometimes said that the social sciences lag behind the natural because controlled experiments cannot be done so readily in the former. Moreover in certain sciences, e.g. epidemiology, students are explicitly taught that experimental data is preferable to observational data, particularly for doing causal inference. So the distinction between observational and experimental science has quite wide currency, and is often regarded as methodologically significant.

Surprisingly, mainstream philosophy of science has had rather little to say about the observational/experimental distinction.² For example, discussions of confirmation usually invoke a notion of 'evidence', to be contrasted with 'theory' or 'hypothesis'; the aim is to understand how the evidence bears on the hypothesis. But whether this 'evidence' comes from observation or experiment generally plays no role in the discussion; this is true of both traditional and modern confirmation theories, Bayesian and non-Bayesian. The same is true in the realism/anti-realism debate,

¹ On this point, see the insightful discussion of L.J. Savage (1954), p.117-8, which concludes "however useful the distinction between observation and experiment may be in ordinary practice, I do not yet see that it admits of any solid analysis" (p.118).

² There has been considerable philosophical work on scientific experimentation, e.g. Franklin (1986), (1990), Galison (1987), Weber (2005), but the contrast with observation does not feature prominently in this literature.

where the notion of ‘underdetermination of theory by data’ features prominently; whether the ‘data’ are observational or experimental rarely even gets a mention.³

This is surely a lacuna. If the observational/experimental distinction has the methodological significance often imputed to it, one might look naturally to philosophy of science to explain this. In this paper I want to sketch one possible explanation, by suggesting that observation and experiment will often differ in their confirmatory power. Based on a simple Bayesian analysis of confirmation, I argue that universal generalizations (or ‘laws’) are typically easier to confirm by experimental intervention than by pure observation. This is not to say that observational confirmation of a law is impossible, which would be flatly untrue – think of how Kepler confirmed his laws of planetary motion. But there is a general reason why confirmation will accrue more easily from experimental data, based on a simple though oft-neglected feature of Bayesian conditionalization.

I want to approach the issue in a roundabout way, by discussing an apparently unrelated question, namely whether there is a Bayesian vindication of the traditional hypothetico-deductive (H-D) principle. By the H-D principle I mean the principle that scientific theories are confirmed when they make correct predictions. The H-D principle has been criticised, but is intuitively compelling and fits many real scientific cases. Many Bayesians claim that their theory can explain why the H-D principle is valid when it is, and so provides a sort of justification for it.⁴ The explanation goes like this. Suppose that a hypothesis H logically implies a testable statement e ; so $P(e/H) = 1$. Suppose that initially we don’t know whether e is true but consider it possible; so $0 < P(e) < 1$. Similarly, we don’t know whether H is true but consider it possible; so $0 < P(H) < 1$. Under these conditions, Bayes’ theorem implies that $P(H/e) > P(H)$, i.e. e confirms H . If we then learn that e is true and update by Bayesian conditionalization, our credence in H must increase. Thus we have an explanation of the kernel of truth in the H-D principle. John Earman (1992) describes this as one of the main ‘success stories’ of Bayesian confirmation theory.

This ‘success story’ could be criticised on the grounds that there is only rarely a strict deductive link between hypothesis and prediction, as Duhem (1906) famously argued. That is no doubt true; but the ‘ H implies e ’ model might still be a reasonable

³ An exception is the ‘entity realism’ of Hacking (1983), which accords special epistemic status to entities that can be experimentally manipulated.

⁴ This claim is made by Earman (1992), Howson and Urbach (1993), and Strevens (2006) among others.

idealization, particularly if the auxiliary assumptions needed to make the link truly deductive are much more certain than the hypothesis itself. However there is another reason for doubting the simple Bayesian explanation of the H-D principle. Very often, what we learn from experience will be *logically stronger than what our hypothesis implies* (cf. van Fraassen 1986 p.284). Grant that ‘H implies e’ is a reasonable representation of the link between a scientific hypothesis and an empirical prediction. Certainly, if a scientist learns that e is true *and learns nothing else*, updating by conditionalization will lead him to become more confident of H (presuming both e and H have non-extremal priors). But what if e is only *part* of what the scientist learns? In that case, updating on the new information need not raise H’s probability and could lower it – even though this information contains e as a part.

The general point here is that Bayesian conditioning requires one to update on the logically strongest proposition one learns, i.e. the totality of information received.⁵ So if the hypothesis H implies e, but the strongest proposition the scientist learns is e’ (where $e' \Rightarrow e$), the scientist must conditionalize on e’ to conform to Bayesian principles. If she does so, there is no guarantee that she will raise her probability of H, since e’ is not a logical consequence of H. So if the scientist does end up more confident of H after learning e’, i.e. does regard H as confirmed, there is no automatic Bayesian explanation of this. The scientist’s epistemic shift may be compatible with Bayesian updating, but not necessarily.

When we learn information about the world, in a scientific or everyday context, it may be hard to say what the logically strongest proposition learnt is; one might even doubt whether there always *is* such a proposition, in every learning episode. However this may be, it should not blind us to the fact that there is such a thing as throwing information away. If I hear on the radio that there has been an earthquake in Japan, then the proposition that there has been an earthquake in Asia is only part of what I have learnt. This simple point is all that we need here.

⁵ This point is sometimes described as ‘the principle of total evidence’. This is unfortunate, both because Carnap originally used that label for a different idea, and because it makes the point sound like a methodological maxim about which there could be some debate. But in fact, conditioning on only part of the total information received simply leads to inconsistency. Suppose e’ is the logically strongest proposition learnt, and $e' \Rightarrow e$. Suppose an agent conditionalizes on e, i.e. obeys the rule $P_{\text{new}}(x) = P_{\text{old}}(x/e)$ for all x in the domain of P. What is the value of $P_{\text{new}}(e')$? Conditionalization implies that the answer is $P_{\text{old}}(e'/e)$, but the fact that the agent has learnt e’ implies that the answer is 1, which is a contradiction.

Often when a scientist puts a hypothesis to empirical test and the hypothesis passes, the information learnt will be logically stronger than what the hypothesis implies. The hypothesis that humans and chimps share a recent common ancestor implies (*modulo* certain auxiliaries) that the two species should be genetically similar; what scientists learnt when they examined the human and chimp genomes was that they exhibit a certain *very specific* pattern of similarities. More generally, in many sciences a typical hypothesis might predict that an observable parameter will have a value within a certain bound, e.g. $[0.3, 0.5]$; what scientists learn, when they perform the relevant test, is that the parameter has a particular point-value within that bound, e.g. 0.45. So again, the information received is logically stronger than what the hypothesis implies.

To help see why this undermines the standard Bayesian explanation of the H-D principle, consider the two figures below. Both depict Venn diagrams in which a hypothesis H entails a prediction e . If the scientist learns that e is true, and learns nothing logically stronger, then conditioning on e must lead the scientist to become more confident of H . This is clear from Figure 1: the Venn diagram corresponding to H is a smaller fraction of the total space that it is of the Venn diagram corresponding to e ; so learning that the actual world is inside e must make H more probable. But if what the scientist learns is that e' is true, where $e' \Rightarrow e$, we need to look at Figure 2. In this case, conditioning on the information received, i.e. e' , will not necessarily make H more probable, for H is not nested within e' .

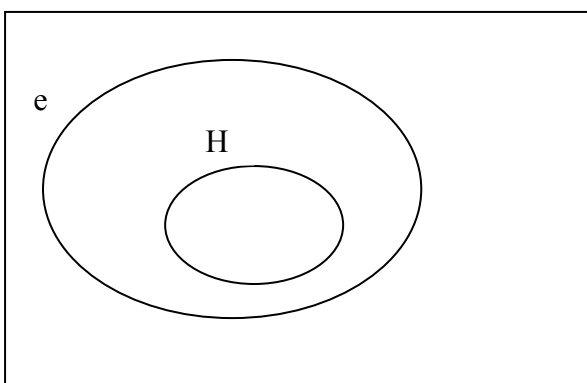


Figure 1

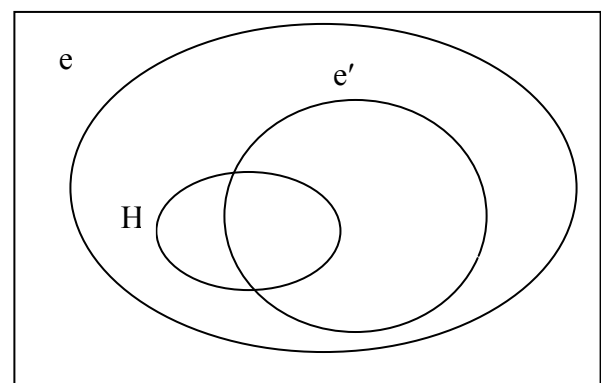


Figure 2

A simple example due to Rosenkrantz (1982) neatly illustrates the importance of updating on all, not just part, of the information one receives.⁶ Five gentlemen, numbered 1 to 4, are at a party, and have left their hats at the door. Consider the hypothesis H that each gentleman leaves the party with someone else's hat. This hypothesis makes a number of testable predictions, including that gentleman 1 leaves with someone else's hat (e_1), that gentleman 2 leaves with someone else's hat (e_2), and so on. Suppose we learn that e_1 , e_2 and e_3 are all true, *and learn nothing else*. This will lead us to become more confident of H , once we update. But suppose we learn the more specific information that gentleman 1 has left with 2's hat (e'_1), that gentleman 2 has left with 3's hat (e'_2), and that gentleman 3 has left with 1's hat (e'_3). This more specific information *conclusively refutes* H – for it implies that gentleman 4 has left with his own hat. So although we have learnt that three predictions of our hypothesis are true (since $e'_1 \Rightarrow e_1$, $e'_2 \Rightarrow e_2$ and $e'_3 \Rightarrow e_3$), this is only *part* of what we have learnt; and the additional information makes a big difference.

What has this got to do with the distinction between observational and experimental science? To see its relevance, consider a universal generalisation of the form 'all F s are G ', i.e. $\forall x(Fx \rightarrow Gx)$. Many philosophers have thought that such generalisations are confirmed by their 'positive instances', i.e. by discovering objects that are both F and G . But we have to consider carefully the informational content learnt from discovering such an object. In some cases, it might plausibly be represented as $[Fa \ \& \ Ga]$, or perhaps as $\exists x(Fx \ \& \ Gx)$. But neither of these is a logical consequence of the universal generalisation. Certainly, $[Fa \ \& \ Ga]$ implies the material conditional $[Fa \rightarrow Ga]$, i.e. that if object a is F then it is G , which *is* a consequence of $\forall x(Fx \rightarrow Gx)$; but as stressed above, one must always condition on the strongest proposition learnt. And conditioning on $[Fa \ \& \ Ga]$ will not necessarily raise the probability of the generalization.

In a purely observational science, where experimental interventions are not possible, $[Fa \ \& \ Ga]$ seems like a plausible representation of the information that might be gained from observation. Experiments being impossible, direct empirical evidence for the generalization can only come from happening upon objects that possess both properties; the informational content of such a 'happening' is plausibly

⁶ The example as described here is a simplified version of Rosenkrantz's original, but makes the same point.

that a certain object possesses both properties, or something reasonably close to this. The important feature is that this involves learning that a given object has property G, where one did not know in advance that the object had property F.

But in an experimental science, where the world can be deliberately manipulated, things are different. In this case, one can gain evidence for the generalization $\forall x(Fx \rightarrow Gx)$ by bringing it about experimentally that a particular object has property F, and then testing to see if it also has G. In this case, 'Fa' is part of one's background knowledge *before* one learns whether Ga or \neg Ga. If one then learns that Ga and conditionalizes, the probability of the generalization *must* increase – for Ga is a logical consequence of the conjunction of Fa and $\forall x(Fx \rightarrow Gx)$, and Fa has prior probability of 1.

In short, there is a crucial difference between discovering that some object is both F and G, and discovering of some object that one *antecedently knows* to be an F, that it is a G.⁷ The latter discovery must always confirm the generalization $\forall x(Fx \rightarrow Gx)$, when one updates by Bayesian conditioning, but the former need not. I suggest that discoveries of the former sort are typical in purely observational sciences, while those of the latter sort typify the experimental sciences. If this is correct, it follows that confirmation of universal laws accrues more easily in the experimental sciences. This in turn provides a partial vindication of the widespread notion that the difference between the two sorts of science is methodologically significant.

An example may help illustrate. Consider the generalization 'impure samples of water have a boiling point above 100C'. An experimental test of this might involve getting a pure sample of water, adding impurities, then seeing if it boils at a temperature above 100C. If it does then the experimenter has learnt, of an object antecedently known to be an impure sample of water, that its boiling point is above 100C. So the generalization is confirmed. Contrast the generalization 'all meteorites landing on earth have diameter greater than 5cm'. Experimental test of this is impossible: one cannot bring it about that a meteorite lands on earth, as one can bring it about that a sample of water is impure. Observation is all there is to go on. So empirical confirmation of the generalization does not involve learning of an object that one has engineered to have property F, that it has property G. More likely, it simply involves coming across an object that is both a meteorite and has diameter

⁷ This distinction is emphasised by Rosenkrantz (1982) and more recently by Fitelson and Hawthorne (forthcoming) in their instructive analysis of Hempel's 'paradox of confirmation'.

greater than 5cm. In this case, the information received is of the form ‘Fa and Ga’, which does not necessarily confirm the generalization.

One might object that even in the experimental case, the information provided by the experiment may still be logically stronger than what the hypothesis entails. In the water example, the experimenter might learn that his impure water sample boils at 103.5C, which implies but is not implied by its boiling point being above 100C. So strictly speaking, Bayesian updating need not lead the experimenter to become more confident of the generalization. This point is correct, but it does not undermine the contrast between the experimental and observational cases altogether. Rather, it highlights that even in experimental cases, it may be an idealization to regard the scientists as learning a logical consequence of his hypothesis. But it is much *less* of an idealization than in observational cases, where it is a serious distortion to treat the information learnt as a consequence of the hypothesis. Where ‘Fa’ is part of background knowledge, as in an experimental set-up, then the idealization may approximately or exactly apply; but where ‘Fa’ is not antecedently known, the scientist who learns that object a is both F and G has *of necessity* learnt more than what the generalization implies, so the idealization has no chance of being correct.

Another possible objection is that the distinction between learning of an object, antecedently known to be F, that it is G, and learning that an object is both F and G, has been overplayed. After all, it is well-known that Bayesian conditionalization is insensitive to order: if one conditionalizes a probability function on x, and then conditionalizes the result on y, one ends up with the same posterior probability function as if one had conditionalized on the conjunction ‘x and y’ in one go. So how can the distinction in question matter?

Two things can be said in response. Certainly, if one conditionalizes on ‘Fa and Ga’ in one go, the posterior probability of the generalization $\forall x(Fx \rightarrow Gx)$ will be the same as if one had updated twice, firstly on Fa and then on Ga. But in the two-step case, the second update has got to be a confirmation, i.e. the probability of the generalization must go up (from what it was after the first update). In the one-step case, no confirmation need occur. So there is still a qualitative difference between the two cases, with respect to how easy it is for the generalization to receive confirmation, even though its final posterior probability must be the same in both cases.

This may seem like a weak response. But there is a deeper reply. In the experimental case, where one intervenes to bring it about that Fa (before testing to see whether Ga), this ‘intervention’ isn’t usefully regarded as an instance of Bayesian learning. By making it the case that Fa (adding impurity to a water sample, in the above example), the scientists obviously adds ‘ Fa ’ to their stock of background knowledge, but this isn’t representable as Bayesian updating. For prior to the experimental manipulation, which made it that case that object a had property F , the scientist knew that the object did *not* have F . This is precisely the sort of epistemic shift that cannot be captured in Bayesian terms; for the proposition learnt is the negation of something that had prior probability one. Not every augmentation of our background knowledge can be modelled as Bayesian updating.

It follows that the objection above, from the order-invariance of Bayesian updating, misses its mark. Suppose my characterization of experimentation versus observation is correct – learning of an object, antecedently known to be F , that it is G versus learning that an object is both F and G . So in experimentation, the scientist first learns F and later learns G ; but it does not follow that ‘learning F ’ must be modelled as Bayesian conditionalizing, and we have seen that it cannot. So even granted my characterization, the difference between experimentation and observation does not equate to the difference between sequential conditionalizing (first on Fa then on Ga) and single-step conditionalizing on their conjunction. So the fact that sequential and single-step conditionalizing would lead to the same posterior probability function is not to the point.

A final objection is this. Surely it is not the fact of experimental intervention itself which matters, but the epistemic situation into which one is thereby put. Suppose a scientist, seeking confirmation of ‘All F s are G ’, experimentally manipulates the world to bring it about that Fa . He thereby puts himself in an epistemic situation whereby if he learns that Ga (and nothing stronger), conditionalization will raise the probability of the generalization. It is being in this epistemic situation that matters, not experimental intervention per se.

This objection is correct, but it does not invalidate the forgoing argument. Rather it forces a re-think of what we mean by ‘observation’ and ‘experiment’. I have characterized observation as ‘happening upon an object that is both F and G ’, but this is a bit naïve. Palaeontology is presumably an observational science, but palaeontologists do not just stumble around until they happen upon fossils; rather they

undertake carefully controlled searches, in specially chosen locations, which may involve deliberate modification of the environment. Such ‘organized’ observation is perhaps more akin to experimentation than observation characterised in the naïve way above. Moreover, ‘organized’ observation *may* be a way of getting into the epistemic situation characterized in the previous paragraph. To see this, suppose an ornithologist is seeking confirmation of ‘all ravens are black’. As we know, such confirmation will accrue if the ornithologist discovers, of some object antecedently known to be a raven, that it is black. Now suppose the ornithologist goes to a bird sanctuary known to contain only ravens, and studies the colour of the birds he finds there. By going to the sanctuary, he has put himself in an epistemic situation whereby if he discovers that some bird is black, he will become more confident of the generalization. Importantly, he needed to *get himself into this epistemic situation* by going to the raven sanctuary to make his observations; he was not in it automatically. Had he merely looked at the birds in some forest, he would not have been in it.

Rather than taking this to show that experimentation is not the only way of getting into the epistemic situation in question, I suggest that we instead conclude that the ornithologist did indeed perform an experiment, of a rudimentary sort. More generally, getting oneself into the epistemic situation in question – in which learning that Ga will lead one to become more confident of $\forall x(Fx \rightarrow Gx)$ – can be regarded as an operational definition of what it is to perform an experimental test of the generalization.

This yields an understanding of ‘experimentation’ (or ‘experimental test’) that may seem rather liberal. However the experimental/observational distinction has never been made precise; and our operational definition may actually explain certain facets of scientific usage. Take for example Eddington’s famous 1919 expedition to test Einstein’s general relativity theory. Einstein had predicted that starlight would be deflected by the sun’s gravitational field, an effect that only would be detectable during a solar eclipse. Eddington travelled to the island of Principe to take measurements during the solar eclipse of May 1919 and found them to be in accord with Einstein’s prediction.

Should we describe what Eddington did as an observation or an experiment? Since Eddington obviously did not modify the heavens to bring starlight within the sun’s gravitational field, ‘observation’ may seem like the right answer; indeed the

word ‘observation’ features in the title of his original paper (cf. Dyson et. al. 1920). However, Eddington’s test is today referred to as one of the three ‘experimental’ tests of general relativity theory, despite the absence of any experimental intervention in the ordinary sense. This is neatly explained using my characterization of ‘experimental test’. By undertaking his expedition, Eddington put himself in an epistemic situation whereby if he made certain observations, his confidence in Einstein’s theory would be boosted. Has the same telescopic observations been recorded but *not* during a solar eclipse, Einstein’s theory would not have been confirmed. In that sense, Eddington did an experiment.⁸

This operational characterization of an ‘experimental test’, via the epistemic position that it puts the scientist in, implies that experiments can occur even when we would be reluctant to say there has been any ‘manipulation’ or ‘intervention’ in the course of nature. But this does not make it trivial. For the key point is that one needs to deliberately get into such an epistemic situation; one is not in it automatically. One good way of getting into the epistemic situation in question – in which were one to learn that Ga one’s confidence in $\forall x(Fx \rightarrow Gx)$ would increase – is to bring it about that Fa by intervention, i.e. to do an experiment in the ordinary sense; but there may be other ways too. And once the scientist has got into that epistemic situation, the Bayesian analysis above shows that confirmation of the generalization will then accrue more easily.

If my characterization of the difference between experimentation and observation is accepted, this goes part way to explaining the common opinions that (i) there is an important methodological difference between the two types of science, and (ii) the former is somehow superior. My argument suggests that there is a kernel of truth in these opinions, for there is a quite general explanation, based on simple Bayesian principles, for why experimentation will more easily permit confirmation of general laws than observation alone can.

Why has this distinctive feature of experimentation not been noted before? The answer, I suspect, is that the notion of a ‘positive instance’ has occupied centre-stage in philosophical discussions of the confirmation of scientific generalizations – a tradition started by Hempel (1945) and continuing today. A ‘positive instance’ of ‘All

⁸ It is worth noting that although astronomy is often cited as a paradigmatic observational science, there is a Springer journal called *Experimental Astronomy*, which “is a medium for the publication of papers of contemporary scientific interest on astrophysical instrumentation and methods necessary for the conduct of astronomy at all wavelength fields”.

Fs are G' is usually described as an *object* that is both F and G; much ink has been spilled on the question of whether 'positive instances' always confirm generalizations, and if not why not. But as Rosenkrantz (1982) and Good (1967) astutely observed, from a Bayesian perspective the notion of 'positive instance' is rather unhelpful. For a Bayesian, the relata of the confirmation relation are propositions, not 'objects'; so we have to ask what exact propositional content is provided by the discovery of a 'positive instance'.⁹ Asking this leads naturally to an appreciation of the crucial distinction between learning that some object is both F and G, and learning of an object, antecedently known to be F, that it is G. This distinction, central to my argument above, remains obscure so long as the notion of 'positive instance' is taken as primitive.

⁹ See Fitelson (2008) for further discussion of the problems that arise from treating 'objects' rather than propositions as the first relatum of the confirmation relation.

References

- Duhem, P. (1908) *The Aim and Structure of Physical Theory*, Princeton: Princeton University Press.
- Dyson, F. W.; Eddington, A. S., Davidson C. (1920). "A determination of the deflection of light by the Sun's gravitational field, from observations made at the total eclipse of 29 May 1919". *Philosophical Transactions of the Royal Society of London* 220A: 291–333.
- Earman, J. (1992) *Bayes or Bust?*, Cambridge MA: MIT Press.
- Fitelson, B. (2008) 'Goodman's New Riddle', *Journal of Philosophical Logic* 37, 613-643.
- Fitelson, B. and Hawthorne, J. (forthcoming) 'The Wason Task(s) and the Paradox of Confirmation', *Philosophical Perspectives* (forthcoming)
- Franklin, A. (1986) *The Neglect of Experiment*, Cambridge: CUP.
- Franklin, A. (1990) *Experiment: Right or Wrong*, Cambridge: CUP.
- Galison, P. (1987) *How Experiments End*, Chicago: Chicago University Press.
- Good, I. J. (1967) 'The White Shoe is a Red Herring', *British Journal for the Philosophy of Science* 17, 322.
- Hacking, I. (1983) *Representing and Intervening*, Cambridge: Cambridge University Press.
- Hempel, C. (1945) 'Studies in the Logic of Confirmation', *Mind* 54, 1-26.
- Howson, C. and Urbach, P. (1993) *Scientific Reasoning: the Bayesian Approach*, 2nd edition, La Salle: Open Court.
- Rosenkrantz, R. (1982) 'Does the Philosophy of Induction Rest on a Mistake', *Journal of Philosophy* 79, 78-97.
- Savage, L. J. (1954) *The Foundations of Statistics*, New York: Wiley.
- Strevens, M. (2006) 'Notes on Bayesian Confirmation Theory', unpublished m/s.
- van Fraassen, B. (1985), 'Empiricism in the Philosophy of Science', in P. Churchland and C.A. Hooker (eds.) *Images of Science*, Chicago: University of Chicago Press, 245-308.
- Weber, M. (2005) *Philosophy of Experimental Biology*, Cambridge: CUP.